

Polarization as a novel architecture to boost the classical mismatched capacity of B-DMCs

Mine Alsan and Emre Telatar
EPFL – I&C – LTHI
Lausanne, Switzerland
Email: {mine.alsan,emre.telatar}@epfl.ch

Abstract—We show that the mismatched capacity of binary discrete memoryless channels can be improved by channel combining and splitting via Arıkan’s polar transformations. We also show that the improvement is possible even if the transformed channels are decoded with a mismatched polar decoder.

Index Terms—Mismatched capacity, mismatched polar decoder, polar transforms

I. INTRODUCTION

In various communication scenarios, we encounter sub-optimal decoders due to practical implementation constraints (complexity, feasibility requirements) or partial/missing channel information. In general, these decoders might perform worse than optimal decoders which minimize the average decoding error probability and possibly result in capacity loss. Modeling such sub-optimal scenarios via ‘reliable communication with a given decision rule’ and establishing coding theorems for them allows one to assess the extent of any loss.

To allow their study within a unified framework, decoders can be categorized based on generic definitions of their decision functions. Csiszár and Narayan [1] studied and surveyed the performance of those using additive decision rules: Given a codeword set $\{\mathbf{x}(1), \dots, \mathbf{x}(M)\} \in \mathcal{X}^n$, an (additive) d -decoder assigns a received sequence \mathbf{y} , the message $i = 1, \dots, M$ such that $d(\mathbf{x}(i), \mathbf{y}) < d(\mathbf{x}(j), \mathbf{y})$ for $\forall j \neq i$. (If there is no such i the decoder declares an erasure). The decision function is computed using the additive extension of a single letter metric $d(x, y)$. Note that the optimal maximum likelihood (ML) decoding rule is included in this class. In addition, a family of decoders of practical interest within the class correspond to mismatch decoders. In this case, the usual ML rule is kept, but instead of the true channel law a different one is employed in the decision procedure.

The transmission capacity of the channel W when decoded with an additive metric d is denoted by $C_d(W)$. When the metric d corresponds to the ML decoder with respect to a channel V , we denote the corresponding capacity by $C(W, V)$. No closed form single letter expression is known for $C(W, V)$ or $C_d(W)$. Single-letter lower bounds have been derived, but no converse for any of the lower bounds exists, except for some special cases. Binary input binary output channels are such a case where $C_d(W) = C(W)$ or 0 depending on whether or not the mismatch metric is in ‘harmony’ with the channel behavior [1]. Another exception

is the class of binary input discrete memoryless channels (B-DMC). Balakirsky [2] gave a computable expression for $C_d(W)$ when W is a B-DMC. See [1], for a more detailed survey on the topic and a more complete list of references.

One of the recent breakthroughs in coding theory is Arıkan’s invention of polar codes [3]. A polar code of blocklength N is specified by selecting an information set \mathcal{A} over which information is to be transmitted and by freezing the remaining inputs to known values. More specifically, one starts by polarizing the channel by the recursive application of the basic polarization transformations defined by Arıkan. The repeated application yields at stage n , a set of $N = 2^n$ synthetic channels $\{W_N^{(i)} : i = 1, \dots, N\}$. Then, the information set is constructed by selecting the channel indices which are good for uncoded data transmission. Once the data u_1^N is transmitted, the SCD of polar codes will decode the outputs y_1^N using the following estimators [3]

$$\hat{u}_i = \begin{cases} u_i, & \text{if } i \in \mathcal{A}^c \\ f^{(i)}(y_1^N, \hat{u}_1^{i-1}), & \text{if } i \in \mathcal{A} \end{cases}, \quad (1)$$

where Arıkan uses for $f^{(i)}(y_1^N, \hat{u}_1^{i-1})$ the ML decoding rule over the i -th synthetic channel $W_N^{(i)}$.

In fact, SCDs can be considered as another large family based on successive cancellation decoding procedures. Offering a quite different decoding paradigm than additive decoders, a given SCD will decode the received output \mathbf{y} in N stages using a chain of estimators from $i = 1, \dots, N$ each possibly depending on the previous ones. The estimators \hat{u}_i can base their decisions on arbitrary single letter metrics of the form $d_i(u_i, y_1^N \hat{u}_1^{i-1})$. The SCD of polar codes, however, owes its fame not only for yielding polar coding theorems proving the ‘symmetric capacity achievingness’ of polar codes for a large class of channels, but also for inheriting the low complexity structure of the recursive code construction process.

In [4] and [5], the performance of polar codes over an unknown B-DMC with mismatched successive cancellation decoding is investigated. The mismatched polar decoder operates according to decision functions $f^{(i)}(y_1^N, \hat{u}_1^{i-1})$ corresponding to the ML rule of mismatched channels $V_N^{(i)}$ synthesized by polarizing a B-DMC V different than the true communication channel W . A lower bound denoted by $I(W, V)^1$ is obtained in

¹See Eq. (18) for the mathematical definition

[5] on the achievable rates by mismatched polar codes over any B-DMC; $I(W, V)$ turns out to be, as the symmetric capacity of the channel [3], a conserved quantity under the polar transforms. While preserving the low complexity structure of the original polar decoder, the mismatched polar decoder extends the theory of channel polarization and polar codes to mismatched processes.

Let $C_P(W, V)$ denote the mismatched capacity of a polar code transmitted over the channel W when decoded with a mismatched polar SCD designed with respect to another channel V . The lower bound $I(W, V)$ derived in [5] for $C_P(W, V)$ of B-DMCs is also a lower bound to $C(W, V)$, see Fischer [6]. On the other hand, no general order between the polar mismatched capacity $C_P(W, V)$ and the classical mismatched capacity $C(W, V)$ can be formulated: there are examples for which $C_P(W, V) > C(W, V)$ and also examples where the reverse inequality holds.

The main goal of this paper is to show that:

(g1) There are pairs of B-DMCs W and V for which

$$C(W^+, V^+) + C(W^-, V^-) \geq 2C(W, V)$$

holds.

(g2) Furthermore, there exist cases for which

$$C_P(W, V) > C(W, V)$$

holds.

For that purpose, we shall study the evolution of the mismatched capacity of B-DMCs under the one-step polarization transformations when the communication channel and the mismatched channel used in the decision procedure are both symmetrized by the same permutation, i.e., when for some permutation π on the output alphabet satisfying $\pi = \pi^{-1}$, we simultaneously have $W(y|0) = W(\pi(y)|1)$ and $V(y|0) = V(\pi(y)|1)$ for all output letters y .

The rest of this paper is organized as follows. Section II will briefly introduce the necessary definitions and existing results. Then, some analytical computations and numerical experiments will be presented in the Section III. Finally, the paper will close with some discussions.

II. PRELIMINARIES

We start by presenting the results of [2] on the classical mismatched capacity of B-DMCs.

Definition 1: [1] The d -capacity of a discrete memoryless channel W , denoted by $C_d(W)$, is defined as the supremum of rates R for which, for every $\epsilon > 0$ and sufficiently large block length n , there exist codes with d -decoding such that the code rate $\frac{\log M}{n} > R$, and the maximum probability of error is less than ϵ .

Let $W : \mathcal{X} \rightarrow \mathcal{Y}$ be a B-DMC with $\mathcal{X} = \{0, 1\}$. We fix an input distribution $P(x)$ on \mathcal{X} and denote the transition probabilities of the channel by $W(y|x)$. Some standard definitions

follows:

$$H(PW) = - \sum_{y \in \mathcal{Y}} PW(y) \log_2 PW(y), \quad (2)$$

$$\text{with } PW(y) = \sum_{x \in \mathcal{X}} P(x)W(y|x).$$

$$H(W|P) = - \sum_{y \in \mathcal{Y}} P(x)W(y|x) \log_2 W(y|x). \quad (3)$$

$$I(P, W) = H(PW) - H(W|P). \quad (4)$$

The following result due to Balakirsky gives a closed form expression for $C_d(W)$ when W is a B-DMC.

Theorem 1: [2] For any B-DMC W and any additive decoding metric $d(x, y)$

$$C_d(W) = \max_P I_d(P, W), \quad (5)$$

where

$$I_d(P, W) \triangleq \min_{W': \substack{PW'(y)=PW(y) \\ d(P, W') \leq d(P, W)}} I(P, W'), \quad (6)$$

$$\text{with } d(P, W) = \sum_{x, y} P(x)W(y|x)d(x, y).$$

Using this closed form expression, Balakirsky studies the computation of $C_d(W)$ for symmetric B-DMCs when the d -decoder preserves the symmetry structure of the communication channel. In the following two examples, we revisit his examples.

Example 1: [2, Examples, Statement 1] For a binary input binary output channel W and any given d -decoding rule $I_d(P, W) = I(P, W)\mathbf{1}_{\{A\}}$ holds with

$$\begin{aligned} A &= \{\text{sign}(1 - W(0|0) + W(1|1)) \\ &= \text{sign}(d(0, 0) + d(1, 1) - d(0, 1) - d(1, 0))\}. \end{aligned} \quad (7)$$

So, $C_d(W) = C(W)$ or 0.

Example 2: [2, Examples, Statement 2] Let $W : \mathcal{X} \rightarrow \mathcal{Y}$ be a B-DMC with $\mathcal{X} = \{0, 1\}$ and $\mathcal{Y} = \{0, 1, \dots, L-1\}$. Suppose the transition probability matrix of the channel W and the corresponding metrics for the additive d -decoder are given by

$$W = \begin{bmatrix} w_0 & w_1 & \dots & w_{L-1} \\ w_{L-1} & w_{L-2} & \dots & w_0 \end{bmatrix}, \quad (8)$$

and

$$d = \begin{bmatrix} d_0 & d_1 & \dots & d_{L-1} \\ d_{L-1} & d_{L-2} & \dots & d_0 \end{bmatrix}. \quad (9)$$

Then, the mismatched capacity is achieved for P uniform on $\{0, 1\}$ and is given by

$$C_d(W) = H(PW) - H(W'|P), \quad (10)$$

where W' is given by

$$W' = \begin{bmatrix} w'_0 & w'_1 & \dots & w'_{L-1} \\ w'_{L-1} & w'_{L-2} & \dots & w'_0 \end{bmatrix}, \quad (11)$$

with

$$w'_y = (w_y + w_{L-1-y}) \frac{e^{-\alpha d_y}}{e^{-\alpha d_y} + e^{-\alpha d_{L-1-y}}}, \quad (12)$$

for $y \in \mathcal{Y}$, and the parameter $\alpha \geq 0$ is chosen from the condition:

$$\sum_{y \in \mathcal{Y}} w'_y d_y = \sum_{y \in \mathcal{Y}} w_y d_y. \quad (13)$$

For the rest of this paper, we will restrict the additive decoders to mismatched decoders. Let $V : \mathcal{X} \rightarrow \mathcal{Y}$ be a B-DMC symmetrized by the same permutation as the channel W defined in Example 2. Suppose the transition probability matrix of V is given by

$$V = \begin{bmatrix} v_0 & v_1 & \dots & v_{L-1} \\ v_{L-1} & v_{L-2} & \dots & v_0 \end{bmatrix}. \quad (14)$$

Then, the corresponding additive decoder can be defined as in (9) by letting $d_y = -\log v_y$, for $y = 0, \dots, L-1$. In this case, the mismatched capacity equals

$$C(W, V) = I(P, W') \quad (15)$$

where the transition probabilities of W' can be computed by replacing the relations in (12) and (13) by

$$w'_j = (w_j + w_{L-1-j}) \frac{\alpha v_j}{\alpha v_j + \alpha v_{L-1-j}}, \quad (16)$$

and

$$\sum_{y \in \mathcal{Y}} w'_y \log v_j = \sum_{y \in \mathcal{Y}} w_y \log v_j. \quad (17)$$

Next, we give the definition of the channel polarization process. From two independent copies of a given binary input channel $W : \mathcal{X} \rightarrow \mathcal{Y}$, the polar transformations synthesize two new binary input channels $W^- : \mathcal{X} \rightarrow \mathcal{Y}^2$ and $W^+ : \mathcal{X} \rightarrow \mathcal{Y}^2 \times \mathbb{F}_2$ with the transition probabilities given by [3]

$$W^-(y_1 y_2 | u_1) = \sum_{u_2 \in \mathbb{F}_2} \frac{1}{2} W(y_1 | u_1 \oplus u_2) W(y_2 | u_2)$$

$$W^+(y_1 y_2 u_1 | u_2) = \frac{1}{2} W(y_1 | u_1 \oplus u_2) W(y_2 | u_2).$$

In analyzing the properties of this channel it is useful to introduce [7] an auxiliary stochastic process, W_0, W_1, \dots , defined by $W_0 := W$, and for $n \geq 0$

$$W_{n+1} := \begin{cases} W_n^- & \text{with probability } 1/2 \\ W_n^+ & \text{with probability } 1/2 \end{cases}$$

with the successive choices taken independently. In this way, W_n is uniformly distributed over the set of 2^n channels above.

Finally, we introduce the results of [5] on the achievable rates by mismatched polar codes. Given two B-DMCs W and V , $I(W, V)$ is defined as

$$I(W, V) = \sum_y \sum_{x \in \{0,1\}} \frac{1}{2} W(y|x) \log_2 \frac{V(y|x)}{\frac{1}{2} V(y|0) + \frac{1}{2} V(y|1)}. \quad (18)$$

Then, we define the process $I(W_n, V_n)$ associated to the polarization process of both channels. Note that $I(W_n) \triangleq I(W_n, W_n)$ corresponds to the symmetric capacity process of the channel. The following result shows that $C_P(W, V) \geq I(W, V)$.

Theorem 2: [5] Let W and V be two B-DMCs such that $I(W, V) > -\infty$. Then, $I_\infty(W, V)$ is $\{0, 1\}$ valued with

$$\mathbb{P}[I_\infty(W, V) = 1] \geq I(W, V). \quad (19)$$

Moreover, the speed of convergence is $O(2^{-\sqrt{N}})$.

III. RESULTS

To begin with, we discuss a trivial improvement of the lower bound stated in Theorem 2. Letting $x^* = \max\{x, 0\}$, the bound in (19) can be improved initially as

$$\mathbb{P}[I_\infty(W, V) = 1] \geq I(W, V)^*. \quad (20)$$

Going a further step, we improve the bound to

$$\mathbb{P}[I_\infty(W, V) = 1] \geq \frac{1}{2} I(W^-, V^-)^* + \frac{1}{2} I(W^+, V^+)^*, \quad (21)$$

and more generally to

$$\mathbb{P}[I_\infty(W, V) = 1] \geq \frac{1}{2^n} \sum_{s \in \{+, -\}^n} I(W^s, V^s)^*, \quad (22)$$

for any $n = 0, 1, \dots$

Now, we study the following example.

Example 3: Let W be a BSC of crossover probability $\epsilon \in [0, 0.5]$ and V be the BSC of crossover probability $1 - \epsilon$. In this example, we will answer the following questions:

- (q1) Suppose we transmit over the channel W , but do mismatch decoding with respect to V as shown in Fig. 1. What is the mismatched capacity $C(W, V)$? What is the value of $I(W, V)^*$?
- (q2) Suppose we first apply the polar transforms to synthesize the channels W^+ , W^- and V^+ , V^- , and then we communicate using the architecture given in Fig. 2. What are the mismatched capacities $C(W^+, V^+)$ and $C(W^-, V^-)$ in this case? What are the values of $I(W^+, V^+)^*$ and $I(W^-, V^-)^*$?
- (q3) Suppose we communicate over the channel W using polar coding, and we do mismatched polar decoding with respect to the channel V . The communication architecture is shown in Fig. 3. What is the mismatched capacity of polar coding $C_P(W, V)$?

Once the answers are derived, we will discuss how this example helps us achieve the two goals we set in the introduction.

- (a1) In this case, the crossover probabilities of the BSCs are not in harmony. By the result given in Example 1, we conclude $C(W, V) = 0$. As $C(W, V) \geq I(W, V)$, we have $I(W, V)^* = 0$.

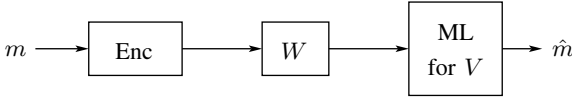


Fig. 1. Classical mismatched decoding.

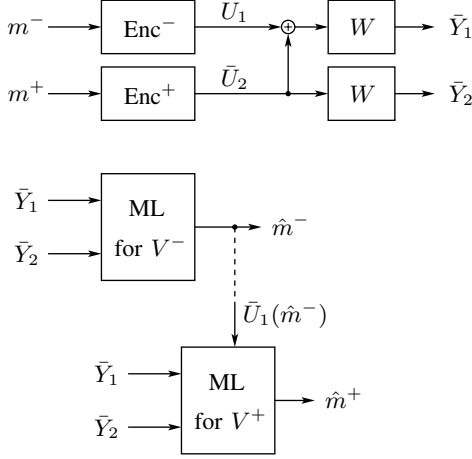


Fig. 2. One-step polarization architecture for mismatched decoding.

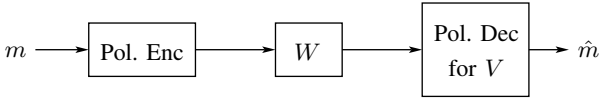


Fig. 3. Polar mismatched decoding.

- (a2) It is known that after applying the minus polar transform to a BSC of crossover probability $\alpha \in [0, 1]$, the synthesized channel is also a BSC, and with crossover probability $2\alpha(1-\alpha)$. So, both W^- and V^- are the same BSC with crossover probability of $\epsilon^- = 2\epsilon(1-\epsilon)$. Therefore, the mismatched capacity of the minus channel equals its matched capacity, i.e., $C(W^-, V^-) = C(W^-) = 1 - h(\epsilon^-)$. Similarly, $I(W^-, V^-)^* = C(W^-) = 1 - h(\epsilon^-)$.
- (a3) We have already seen that $V^- = W^-$. It is easy to see that while $V^+ \neq W^+$, one has $V^{+-} = W^{+-}$, and indeed, $V^{++-} = W^{++-}$, Consequently, for any sequence $s^n = s_1, \dots, s_n$ of polar transforms $V^{s^n} = W^{s^n}$ except when $s^n = +\dots+$. Thus, $I(W^{s^n}, V^{s^n}) = I(W^{s^n})$ for all s^n except $+\dots+$, and we see that $C_P(W, V) = C(W)$, and we arrive at our second goal (g2).

After such a motivating example, one is curious about whether the improvement we illustrated for the specific pair of BSCs extend to other pairs of mismatched B-DMCs as well. To satisfy one's curiosity, we now present the results of the

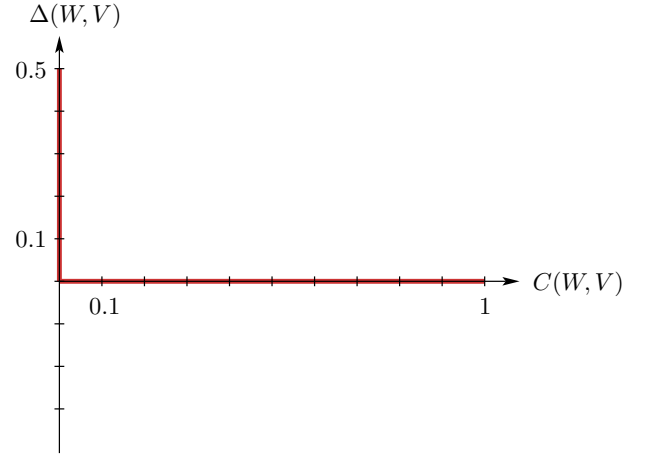


Fig. 4. $C(W, V)$ versus $\Delta(W, V)$ for $|\mathcal{Y}| = \{2, 3\}$.

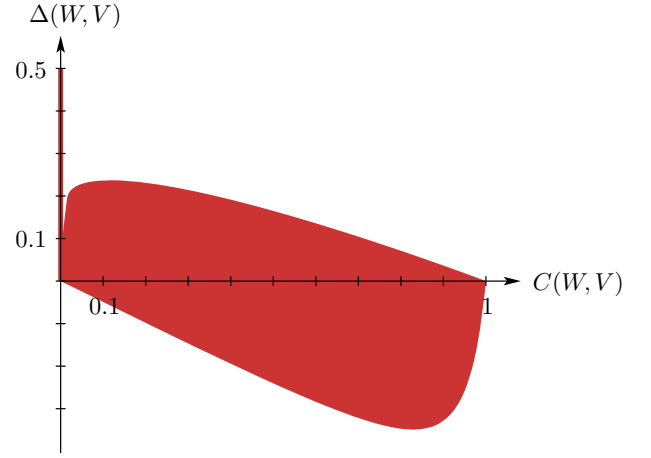


Fig. 5. $C(W, V)$ versus $\Delta(W, V)$ for $|\mathcal{Y}| = 4$.

numerical experiments we carried for random pairs of channels with various output alphabet sizes. Let

$$\Delta(W, V) \triangleq C(W^+, V^+) + C(W^-, V^-) - 2C(W, V). \quad (23)$$

The numerical experiments show that:

- (i) When the output alphabet is binary or ternary, one-step improvement, i.e. $\Delta(W, V) > 0$, happens only when $C(W, V) = 0$ and can be as large as $1/2$. When $C(W, V) > 0$, we observe that $\Delta(W, V) = 0$, and one has neither improvement nor loss. See Fig. 4.
- (ii) When the output alphabet contains four or more symbols, improvement may happen not only when $C(W, V) = 0$ but also when $C(W, V) > 0$; however, there are cases when $C(W^+, V^+) + C(W^-, V^-) < 2C(W, V)$, so one may encounter a loss, i.e. $\Delta(W, V) < 0$, after a one-step transformation. See Fig. 5.

Finally, the numerical experiments also suggest that

$$\sum_{s \in \{+, -\}^n} \frac{1}{2^n} I(W^s, V^s)^* \leq C(W, V) \quad (24)$$

holds whenever $C(W, V) > 0$, (we did experiment only for $n = 1, 2, 3$). Thus, $C_P(W, V) \leq C(W, V)$ is a likely conjecture for the case where $C(W, V) > 0$.

IV. DISCUSSIONS

The study by Balakirsky [2] gave the initial impulse for this work: We adapted his example [2, Examples, Statement 2] which computes $C_d(W)$ of a symmetric B-DMC W when the additive decoder shares the symmetry structure of W to mismatched decoders, and we carried numerical experiments based on this computation to compare $C(W, V)$ with various other quantities such as the sum $C(W^+, V^+) + C(W^-, V^-)$. The experiments reveal that, as opposed to $I(W, V)$, $C(W, V)$ is not necessarily a conserved quantity under the polar transformations. Nevertheless, communication rates higher than $C(W, V)$ can be achieved in some cases by integrating the polarization architecture of Arikan into the classical mismatched communication scenarios. Furthermore, by studying a specific pair of BSCs W and V for which $C(W, V) = 0$, but $C_P(W, V) > 0$, we showed that there exists channels for which the polar transformations strictly improve the mismatched capacity of B-DMCs.

ACKNOWLEDGMENT

This work was supported by Swiss National Science Foundation under grant number 200021-125347/1.

REFERENCES

- [1] I. Csiszar and P. Narayan, "Channel capacity for a given decoding metric," in *Information Theory, 1994. Proceedings., 1994 IEEE International Symposium on*, 1994, pp. 378–.
- [2] V. Balakirsky, "Coding theorem for discrete memoryless channels with given decision rule," in *Algebraic Coding*, ser. Lecture Notes in Computer Science, 1992, vol. 573, pp. 142–150.
- [3] E. Arikan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3051–3073, 2009.
- [4] M. Alsan, "A lower bound on achievable rates over B-DMCs with mismatched polar codes," in *Proc. IEEE Inf. Theory Workshop*, 2013, pp. 6–10.
- [5] —, "Universal polar decoding with channel knowledge at the encoder," Available: *arXiv:1311.7590*, 2013.
- [6] T. R. M. Fischer, "Some remarks on the role of inaccuracy in Shannon's theory of information transmission," in *Trans. of the Eighth Prague Conference*, ser. Czechoslovak Academy of Sciences. Springer Netherlands, 1978, vol. 8A, pp. 211–226.
- [7] E. Arikan and I. Telatar, "On the rate of channel polarization," in *Proc. IEEE Int. Symp. Inf. Theory*, 2009, pp. 1493–1495.